Available online at www.sciencedirect.com
SciVerse ScienceDirect
ELSEVIER
Procedia Computer Science 9 (2012) 1263 – 1272
Procedia Computer Science

International Conference on Computational Science, ICCS 2012

A new application of the support vector regression on the construction of financial conditions index to CPI prediction

Ye Wang[a,b]*, Bo Wang[b,c], Xinyang Zhang[b]

[a]School of Management, Graduate University of Chinese Academy of Sciences, Beijing, China
[b]Research Centre on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, China
[c]School of Mathematical Sciences, Graduate University of Chinese Academy of Sciences, Beijing, China

Abstract

A regression model based on Support Vector Machine is used in constructing Financial Conditions Index (FCI) to explore the link between composite index of financial indicators and future inflation. Compared with the traditional econometric method, our model takes the advantage of the machine learning method to give a more accurate forecast of future CPI in small dataset. In addition, we add more financial indicators including M2 growth rate, growth rate of housing sales and lag CPI in our model which is more in line with economy. A monthly data of Chinese CPI and other financial indicators are adopted to construct FCI (SVRs) with different lag terms. The experiment result shows that FCI (SVRs) performs better than VAR impulse response analysis. As a result, our model based on support vector regression in construction of FCI is appropriate.

Keywords: Financial conditions index, CPI inflation, data mining, support vector regression, financial time series predicting

1. Introduction

Deriving from the study of MCI (monetary conditions index), research about financial conditions index (FCI) has received extensive attention in last ten years. The aim of FCI is supposed to describe the overall financial conditions, and it is a useful indicator for the formulation of monetary policy.

However, econometric research of FCI calculation methods relies on some strong assumptions like linear relationship between inputs variables and output variables. Actually, the relationship of interest rates, exchange rates, fluctuations of asset prices and economic indicators, such as future inflation is not simply linear. Sophisticated financial systems researchers (Jan Hatzius et al. 2010) [1] believed that the transmission channels were diverse and changed over time. Consequently, FCI calculated by econometric method certainly ignores some real economic information.

Support vector machine (SVM) algorithm developed by Vapnik (2000) [2] is based on statistical learning theory. It is a theory of machine learning focus on small sample data based on the structural risk minimization principle from machine learning theory. The algorithm uses a nonlinear mapping from the original data space into some high dimension feature space, and then it constructs a linear discriminate function to replace the nonlinear functions in the original data space. This special character assures that SVM has good generalization ability and it has been widely used in many fields [3][4][5][6][7].SVM is a rather robust tool for forecasting. It is promising methods for prediction of financial time series, Kyoung-jae Kam (2003) [8]. Trafalis et al. (2004) [9] compared the forecasting stock prices of SVM for regression with Back propagation and RBF networks, and found that SVM for regression is a robust technique for function approximation.

Besides solving classification problems, SVMs can also be applied to regression problems by introduction of an alternative loss function, and the loss function must be modified to include a distance measure. Support vector regression (SVR) also can add in the nuclear non-linear regression function which is econometric or other statistic methods lack, so SVR obtains good application of predicting economic indicators and financial time series.

This paper proposes a new method to construct FCI by taking advantage of SVR to forecast CPI. After building the predicting CPI model, we use the weight vectors of SVR model to calculate weights of FCI. In order to verify our method, Chinese financial economic indicators and CPI data are adopted to construct FCI (SVR), and we will test our model in sample and out of sample.

doi:10.1016/j.procs.2012.04.138

This paper is structured as follows: Section 2 describes 3 basic definition of FCI, SVR model and VAR impulse response analysis which will be used. Then, our model based on SVR model is proposed to construct FCI (SVRs), and model test methods are also discussed in the section 3. In the fourth section, the dataset of Chinese CPI and financial indicators is used to carry out experiment, furthermore, in sample and out of sample tests are adopted to verify our model. At last, conclusion is in section 5.

2. Related models and basic definitions

In the early, based on the opinion that interest rate and exchange rate affecting the ultimate goal of monetary policy, Freidman (1953)[10], Bank of Canada built a Monetary Conditions Index constructed by weighted average of short term rate and exchange rate in order to measure the tightness of monetary policy in 1990s. David G Mayes et al. (2001) [11] explored how asset prices, particularly house and stock prices, could provide useful indicators of future outputs and inflation changes. Goodhart et al. (2001) [12] also implied that property and equity prices might also play an important role in the transmission of monetary policy via wealth effect and balance sheet effect, so they extended MCI by adding real property and real share prices to obtain the FCI, and the derived index appears to be a useful predictor of future CPI inflation. The idea was posed as follows.

Definition 1 Formulation of FCI

The definition of FCI is as follows, Goodhart et al (2000):

$$FCI_t = \sum_i w_i (q_{it} - \overline{q_{it}}) \tag{1}$$

where q_{it} is the price of asset i in period t, $\overline{q_{it}}$ is the long-run trend or equilibrium value of the price of asset i in period t, and w_i is the relative weight given to the price of asset i in the FCI. Obviously, the key step of building FCI is to select the appropriate asset q and set w.

Definition 2 VAR impulse response analysis

The weights calculation method of FCI is mainly classified into three types: Large-scale macroeconomic model, IS-Curve-based model and impulse response equation based on VAR. Large-scale macroeconometric simulation is based on economic theory to establish multi-equation model to simulate the overall economic process. But it requires a lot of economic data. As a result, it is not available in all countries. Therefore, this method is seldom used. IS curve reflects the relationship between output gap and interest rates, exchange rates, and other financial variables. The weights in the FCI are determined by the coefficients of variable and their significant in the equation through regression analysis. Mayes et al. (2001) [11] ,Goodhart et al. (2001) [12] used the IS curve to construct European countries FCIs. However, IS curve assumes that assets are exogenous, which is not in line with economic realities.

VAR model assumes every variable in the system as endogenous, and lags all the variables used to construct function. Thus, based on VAR model, impulse response analysis to calculate a unit change of every variable in the impact of inflation; finally, according to the size of impact to determine the weight of the financial variables. Goodhart et al. (2001) [12], Céline Gauthier et al. (2003) [13] evaluated G7 and Canada FCIs by using VAR impulse response functions. Dai et al. (2009) [14] constructed the finance condition index based VECM and made use of FCI to forecast inflation in China. However, VAR or VECM model also exists some limitation. One is that it needs to estimate many parameters so they cannot deal with small sample data. And the other one is that time-series must meet the requirement of stable with same order.

Consider the vector autoregressive model, HH Pesaran (1998) [15]:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \psi W_t + \varepsilon_t,\ \ t = 1, 2, \cdots, T \tag{2}$$

Where $X_t = (X_{1t}, X_{2t}, \cdots, X_{mt})^T$ is an $m \times 1$ vector of endogenous variables, W_t is an $q \times 1$ vector of exogenous variables, and $\{\phi_i, i = 1, 2, \cdots, p\}$ and ψ are $m \times m$ and $m \times q$ coefficient matrices, respectively. ε_t is $m \times 1$ vector of disturbance.

Formulation (2) would follow the assumption of all the roots of $\left| I_m - \sum_{i=1}^{p} \phi_i z^i \right| = 0$ fall outside the unit circle, and X_t would be covariance-stationary, and (2) can be rewritten as the infinite moving average representation without endogenous variables:

$$X_t = \sum_{i=1}^{p} A_i \varepsilon_{t-i},\ \ t = 1, 2, \cdots, T \tag{3}$$

Where A_i is an $m\times m$ coefficient matrices, and we can use OLS to estimate (3).

Based on VAR model, an impulse response function measures the effect of shocks at a given point in time on the future values of variables in a dynamical system. It is a appropriate description is to treat an impulse response as the outcome of a conceptual experiment in which the effect of a hypothetical $m\times 1$ vector of shocks of size $z=(z_1,\cdots,z_m)^T$, hitting the economy at time *t* is compared with a base-line profile at time t+n.

Definition 3 SVR model

In the SVMs, SVR (support vector regression) is used to solve regression problem. In order to set the algorithm, the loss function must be modified to include a distance measure. There are some lost functions in the existing literatures such as the conventional least squares error criterion, Laplacian, Huber and $\varepsilon-insensitive$, Steve R. Gunn (1998) [16]. Generally, SVR chooses the $\varepsilon-insensitive$ loss function (L_ε), Vapnik (2000) [17].

$$L_\varepsilon(f(x),y)=\begin{cases}|f(x)-y|-\varepsilon & if\ |f(x)-y|\ge\varepsilon\\ 0 & otherwise\end{cases} \tag{4}$$

Where ε is a given positive number represent the distance around the regression function $f(x)$, y is actual value. The $\varepsilon-insensitive$ loss function means if the distance of $f(x)$ and y does not exceed ε, there is no loss between $f(x)$ and y. Thus, the region enclosed by the tube is known as "$\varepsilon-insensitive\,zone$", which does not provide any loss value of the objective function.

To illustrate the concept of SVR, a linear regression problem is formulated as follows. Consider a training data set $T=\{(x_1,y_1),\cdots(x_l,y_l)\}$, where x_i is the model inputs, y_i is actual value and represents the corresponding output, and l is total number of data patterns. The objective of the regression analysis is to determine the function, N Deng et al. (2009) [18]:

$$f(x)=(\omega\cdot x)+b,\ \omega\in R^n, b\in R$$

Where, ω is weight vector, b is constant. Based on the $\varepsilon-insensitive$ loss function mentioned above, the weight and constant can be estimated by minimizing the following regularized risk function:

$$\begin{aligned}&\min\ \frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)\\ &s.t.\quad y_i-(\omega\cdot x_i)-b\le\varepsilon+\xi_i\\ &\qquad (\omega\cdot x_i)+b-y_i\le\varepsilon+\xi_i^*\\ &\qquad \xi_i,\xi_i^*\ge 0, for\, i=1,2,\cdots,l\end{aligned} \tag{5}$$

Where, $\frac{1}{2}\|w\|^2$ represent confidence interval, $\xi_i+\xi_i^*$ is slack variable, as the application of the $\varepsilon-insensitive$, and C is the regularization constant used to specify the trade-off between empirical risk and regularization term.

Use Lagrangian multipliers and KKT (Karush-Kuhn-Tucker) conditions to construct dual Lagrangian problem to solve problem (5):

$$\begin{aligned}&\min\quad \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j^*-\alpha_j)K(x_i\cdot x_j)+\varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)-\sum_{i=1}^{l}y_i(\alpha_i^*-\alpha_i)\\ &s.t.\quad \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0\\ &\qquad 0\le\alpha_i,\alpha_i^*\le C,\ i=1,2,\cdots,l\end{aligned} \tag{6}$$

Through solving dual problem (6), we can obtain the optimal solution $\overline{\alpha}^* = (\overline{\alpha}_1, \overline{\alpha}_1^*, \cdots, \overline{\alpha}_l, \overline{\alpha}_l^*)^T$, and $\overline{\omega} = \sum_{i=1}^{l} (\overline{\alpha}_i^* - \overline{\alpha}_i) x_i$.

Hence, the form of linear regression function can be written as:

$$f(x) = \sum_{i=1}^{l} (\alpha_i^* - \overline{\alpha}_i) K(x_i \cdot x) + \overline{b} \tag{7}$$

3. Financial Conditions Index based on SVR

In this section, we first discuss the financial indicators which will be added to our model. Then, based on the linear model which is interpreted in section 2, we propose the non-linear model to predict CPI and calculation of the weights in sample and out of sample model test method will be introduced at last.

Generally, most of FCIs included short-term interest rates, real effective exchange rate, real housing prices and the actual stock price. Some FCIs also include a long-term interest rate or a corporate bond risk premium, such as Beaton et al. (2009) [19] added real long-term interest rates and high-yield bond spreads to construct USA FCI. Michael Holz (2005) [20] used the credit growth indicator, indicator of stock market misalignment to build the EMU FCI; Kimberly Jan Hatzius et al. (2010) [21] chose assets include a broad range of quantitative and survey-based indicators besides interest rates and asset prices.

This article also includes short-term interest rates, real effective exchange rate, real estate prices and stock prices like other existing studies. Besides, Feng et al. (2006) [22] thought that the long-term monetary policy in China took money supply as the intermediate goal and operating objectives. Hence monetary will affect output and inflation, and then monetary supply (M2) should be included in the index FCI. In addition, we add lag of inflation target (CPI) to the model, hoping to use past inflation information to predict future possible situation. Stock market capitalization which was used by Goldman Sachs [23] in its periodic reports issued United States, Europe, China and other countries or regions of FCI is also added in this paper. Moreover, different changes on regional house prices lead to smooth fluctuations in house prices index. Accordingly, we add growth rate of real estate sales, and the index is more volatile. All the financial indicators are listed in table1.

Table 1 variable list

	variable	short		variable	short
1	Lag CPI	LCPI	5	6months short-term lending rates of financial institutions	SIR
2	M2	M2	6	Real estate sales in the month up	RSR
3	RMB real effective exchange rate index	ERI	7	Housing sales price index	HPI
4	HuShen 300 price index return	CSI300	8	Stock market capitalization	SMC

After selecting financial indicators (in table 1), we take them as the input variables of the model, that is, $X_{\text{indicators,i}}=$ ($x_{LCPI,i}$, $x_{M2,i}$, $x_{ERI,i}$, $x_{CSI300,i}$, $x_{SIR,i}$, $x_{RSR,i}$, $x_{HPI,i}$, $x_{SMC,i}$), and the output variable is $Y_{\text{i}} = CPI_{\text{i}}$. Based on (2) in the section 2, we have the definition of loss function:

$$L_\varepsilon(f(X_{\text{indicators}}), CPI) = \begin{cases} |f(X_{\text{indicators}}) - CPI| - \varepsilon & if\ |f(X_{\text{indicators}}) - CPI| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{8}$$

After that, our objective of the regression analysis is to determine the function:

$$f(X_{\text{indicators}}) = (\omega_{SVR} \cdot X_{\text{indicators}}) + b_{SVR},\ \omega_{SVR} \in R^8, b_{SVR} \in R \tag{9}$$

A non-linear mapping linear regression performed can be used to map the data into a high dimensional feature space, in the same manner as the non-linear SVC approach, Steve R.Gun (1998) [16]. To solve nonlinear regression problem in SVR, the inputs are nonlinearly mapped into a high dimensional feature space wherein they are linearly correlated with the output, CJ Lu (2009)[24]. Consquenly, we build the objective function of the dual problem by replacing $(\varphi(x_i) \cdot \varphi(x_j))$ with kernel function $K(x, x^{'})$. In our model, we select the RBF[1] kernel to build the nonlinear regression function:

[1] The formula of RBF: $K(X_{indicators,i}, X_{indicators.j}) = \exp(\frac{-\|X_{indicators,\text{i}} - X_{indicators,j}\|^2}{2\sigma^2})$

$$f(X_{\text{indicators}}) = \sum_{i=1}^{l}(\alpha_i^* - \overline{\alpha}_i)(\varphi(X_{indicators,i}) \cdot \varphi(X_{indicators})) + \overline{b_{SVR}}$$
$$= \sum_{i=1}^{l}(\alpha_i^* - \overline{\alpha}_i)K(X_{indicators,i}, X_{indicators}) + \overline{b_{SVR}} \tag{10}$$

Hence, according to (4) in the section 2, we can build and solve the convex quadratic programming problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j^* - \alpha_j)K(X_{\text{indicators},i} \cdot X_{\text{indicators},j}) + \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{l}CPI_i(\alpha_i^* - \alpha_i) \\ s.t. \quad & \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ & 0 \le \alpha_i, \alpha_i^* \le C,\ i = 1,2,\cdots,l \end{aligned} \tag{11}$$

As a result, we obtain the $\overline{\alpha}^* = (\overline{\alpha}_1, \overline{\alpha}_1^*, \cdots, \overline{\alpha}_l, \overline{\alpha}_l^*)^T$

Through calculate the equation

$$\overline{\omega_{SVR}} = \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)X_{indicators,i}$$

the weight vector $\omega_{SVR} = (\omega_{LCPI}, \omega_{M2}, \omega_{ERI}, \omega_{CSI300}, \omega_{SIR}, \omega_{RSR}, \omega_{HPI}, \omega_{SMC})^T$ is determined. Finally, through the formula:

$$w_i = \omega_i \Big/ \sum_{i=1}^{8}\omega_i \tag{12}$$

where w_i is the weight of indicator i, we can calculate the weights of FCI (SVR).

In order to study the impact of different lag term of financial indicators on future CPI, the lagged one period, lag three periods and lag six periods financial indicators are established as FCI (SVR1), FCI (SVR2), FCI (SVR3) based on (1), respectively.

In sample test

We first find out dynamic correlations of FCI (SVRs) with future inflation in sample to verify that FCI (SVRs) plays an important role in predicting future inflation. CPI is set at a point in time, and we examine the next 10 periods of FCI (SVRs) dynamic correlation with the period of the CPI. If the value is positive (values belongs to [-1,1]), indicating that the future of FCI (SVRs) have dynamic correlation with CPI. Moreover, it also illustrates that FCI (SVRs) lead ahead of CPI.

Then, we introduce the Granger causality test to study the link between CPI and FCI (SVRs) refer to Craig Hiemstra (1994) [25]. Let $F(CPI_t | I_{t-1})$ be the conditional probability distribution of CPI_t, set I_{t-1} consisting of an $L_{CPI}-length$ lagged vector of CPI_t, say, $CPI_{t-L_{CPI}}^{L_{CPI}} = (CPI_{t-L_{CPI}}, CPI_{t-L_{CPI}+1}, \cdots, CPI_{t-1})$, and a $L_{FCI(SVRs)}-length$ lagged vector of $FCI(SVRs)_t$, say

$$FCI(SVRs)_{t-L_{FCI(SVRs)}}^{L_{FCI(SVRs)}} = (FCI(SVRs)_{t-L_{FCI(SVRs)}}, FCI(SVRs)_{t-L_{FCI(SVRs)}+1}, \cdots, FCI(SVRs)_{t-1}).$$

If

$$F(CPI_t | I_{t-1}) = F(CPI_t | I_{t-1} - FCI(SVRs)_{t-L_{FCI(SVRs)}}^{L_{FCI(SVRs)}})),\ t = 1,2,\cdots \tag{13}$$

does not hold, then knowledge of past FCI (SVRs) values helps to predict current and future CPI values, and FCI (SVRs) is said to strictly Granger cause CPI.

Out of sample test by AR model

Based on previous research (Goodhart, 2001 [12]), generally using other methods to build FCI has no ideal result in out of sample forecast. This paper takes Goodhart and Hofmann (2001) [12] method, we estimate AR (Autoregressive) model for predicting CPI rate by lag CPI rate, and compared with joining FCI (SVRs) prediction performance.
We estimate the following formula respectively:

$$\Delta cpi_t = \alpha_1 + \sum_{s=1}^{n} \beta_s \Delta cpi_{t-s} + \varepsilon_t \qquad (14)$$

$$\Delta cpi_t = \alpha_2 + \sum_{s=1}^{n} \beta_s \Delta cpi_{t-s} + \sum_{s=1}^{n} \gamma_s FCI(SVRs)_{t-s} + \varepsilon_t \qquad (15)$$

Where Δcpi_t is CPI rate first-order differential.

We will compare the regression results of (14) and (15) by means of R-squared and Root Mean Squared Error (RMSE), where the R-squared indicates the explanatory power of the CPI rate forecasting, the higher R-squared means the greater explanatory power in predicting; RMSE[2] can evaluate the degree of change data. That is to say, smaller the value is, better accuracy the prediction model describes the experimental data.

4. Experimental results and comparison

The data of indicators are adjusted monthly. CPI and M2 are year on year, equalling to the removal of seasonal trends. The variables in the final model include: CPI [3](year on year), M2 (year on year)[4], RMB real effective exchange rate index[5], CSI 300 index return[6], 6months short-term lending rate of financial institutions[7], real estate sales in the month up[8], housing sales price index[9] and the Shanghai Stock exchange A-share market capitalization[10]. Since none of the above series show long-term stable trend, the original data has good randomness. Variables are listed in table 1, and all the monthly data are from January 2006 to December 2010.

To take advantage of the model proposed in the section 3, we use grid research method to select optimal parameters of C, ξ and σ^2 by LIBSVM-2.9 software system. In according to (12), we obtain the FCI (SVRs) weights showing in the table 2.

Table2 Weights to FCI (SVRs)

	LCPI	M2	ERI	CSI300	RSR	HPI	SIR	SMC
FCI(SVR1)	0.583	0.209	0.212	0.902	-0.767	-0.158	0.609	-0.591
FCI(SVR2)	0.487	-0.103	-0.113	0.123	0.167	0.237	0.102	0.098
FCI(SVR3)	0.751	-0.205	-0.016	-0.096	0.042	0.173	0.171	0.181

Note: FCI (SVR1), FCI (SVR2), FCI (SVR3) corresponding to 1-month lag, 3-months lag and 6months lag FCI.

We use the weights in Table 2 to construct calculation of FCI (SVR1), FCI (SVR2) and FCI (SVR3) by formulation (1) (data is from January 2006 to June 2010). As we can see from Figure 1, the FCI (SVR2) is leading the trend of CPI about six months around. FCI (SVR1) and FCI (SVR3) have the same leading terms with CPI. (See figure in the Appendix).

---

2 $RMSE = \sqrt{\frac{(\sum_{i=1}^{n} \varepsilon_i^2)}{n}}$, Where ε_i represents the different between real value and predicted value of sample i.

3 Data sources from the Chinese National Bureau of Statistics.

4 Data sources from East money network.

5 Data sources from Bank for International Settlements

6 Data sources from China security index company.

7 Data sources from People's Bank of China.

8 Data sources from Chinese National Bureau of Statistics.

9 Data sources from Chinese National Bureau of Statistics.

10 Data sources from Shanghai Stock Exchange.

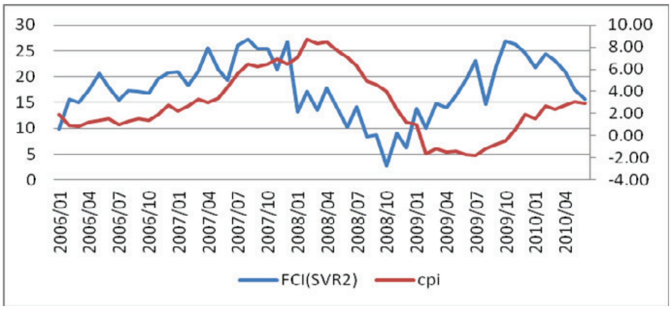

Fig. 1 FCI (SVR2) and CPI

In sample test result

We first look at the result of dynamic correlations between FCI (SVRs) and CPI. Figure 2 shows the dynamic correlations of the above three FCIs with future inflation (Dynamic correlation table in Annex) every month.

As the result, the correlations of FCI (SVRs) with future inflation are generally quite high, show quite a significant correlation between dynamic with CPI. The maximum dynamic correlation coefficients of FCI (SVR1), FCI (SVR2) and FCI (SVR3) were 0.70(5), 0.15(2), 0.60(1), respectively.

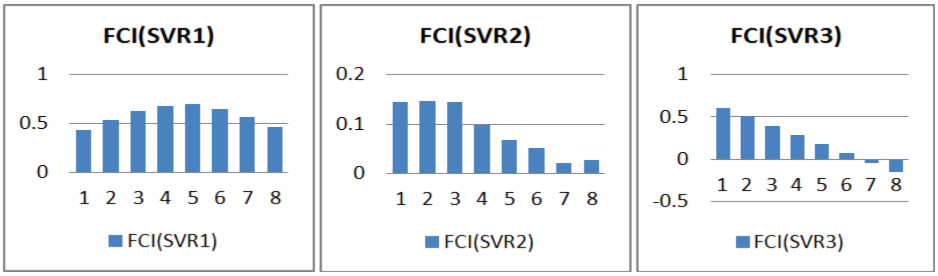

Fig. 2 dynamic correlation results

In addition, we also have Granger causality test described in section3 for in-sample testing. According to the fact that FCI (SVRs) leads 6 months than CPI in above figure 1, 6 lag terms is chosen for further regression. The result shows that the null of no causality is rejected at the 1% level in all cases, so we can conclude that the lagged FCIs contain significant information of future inflation, that is to say, FCI (SVRs) are strictly Granger cause CPI.

Table3 Granger causality test for the FCI (SVRs)

FCI(SVR1)	FCI(SVR2)	FCI(SVR3)
10.474	4.106	42.170

Note: Test statistics are based on a regression including six lagged inflation and FCI (SVRs) terms. The 1% critical value is 3.20.

Out of sample test result

We apply Eviews5.0 software to estimate (14) and (15) to testing FCI (SVRs) in out of sample which is from July 2010 to December 2010. According to regression results, we list R-squared and RMSE in Table 4 and Table 5. It is easy to see, to AR (1) and AR (2), the R-squared added the FCI (SVRs) are larger than the only lag CPI rate model, while, RMSE is less than the only lag CPI rate model. The results show that FCI (SVRs) added variable will effectively improve the accuracy of prediction.

Table4 R-squared and RMSE to first-order autoregressive model

AR(1)	Lag CPI	Lag CPI +FCI(SVR1)	Lag CPI +FCI(SVR2)	Lag CPI +FCI(SVR3)
R-squared	0.798	**0.938**	0.813	0.867
RMSE	0.294	**0.163**	0.284	0.266

Note: "Lag CPI" means that AR model only has CPI rate lags, "Lag CPI+FCI (SVR1)"means that AR model contains CPI rate lags and FCI (SVR1) lags, the same as "Lag CPI+FCI (SVR2)" and "Lag CPI+FCI (SVR3)".

Table5 R-squared and RMSE to second-order autoregressive model

AR(2)	Lag CPI	Lag CPI +FCI(SVR1)	Lag CPI +FCI(SVR2)	Lag CPI +FCI(SVR3)
R-squared	0.926	0.950	0.990	**0.998**
RMSE	0.178	0.146	0.065	**0.027**

Note: "Lag CPI" means that AR model only has CPI rate lags, "Lag CPI+FCI (SVR1)"means that AR model contains CPI rate lags and FCI (SVR1) lags, the same as "Lag CPI+FCI (SVR2)" and "Lag CPI+FCI (SVR3)".

Comparison with VAR

We evaluate FCI (VAR) by VAR impulse response model as previous research generally done to compare with FCI (SVRs). Firstly, go through sequence stationary test because VAR impulse response model requires that all data are stationary with the same order. ADF test results show that the indictors including CPI and RMB real effective exchange rate index at 10% ADF test and PP test significance level cannot reject the null hypothesis, which means that they are non-stationary series. All the above sequence of the ADF test and PP test reject the null hypothesis after respectively first-order differential (the test results see Appendix), which means that there is no unit root, and we use first-order differential for all series to establish VAR model.

VAR model should select first-order lag as the optimal lag through lag order selection based on AIC (Akaike's Information Criterion) and SC (Schwarz Criterion) criteria. While taking advantage of lag 1 data modeling, all modules of unit root within the unit circle, indicating that the VAR model is stable, so the impulse response function can be calculated.

We choose 15 Periods impulse response (impulse response function results in Appendix) in this paper, after then use the $w_i = |z_i| / \sum_{i=1}^{15} |z_i|$ (Goodhart and Hofmann (2001)[12]) to calculate the weights, where w_i is the weight of the variable i, z_i is the average of rate impulse response when the unit Cholesky new innovation of variable i affects CPI inflation in the next 15 months. The weights are listed in table 5, and we can construct FCI (VAR) based on VAR impulse responses by using formula (1).

Table6 Weights to FCI (VAR)

	LCPI	M2	ERI	CSI300	RSR	HPI	SIR	SMC
FCI(VAR)	0.206	0.001	0.134	0.331	0.129	0.138	0.025	0.032

As the same of out-sample test mentioned above, we estimate AR(1) model by adding financial conditions index based on the VAR model and financial conditions index based on the SVRs model, respectively. And then we compare R squared and RMSE. The table reveals that the Financial Conditions Index based on the SVR model performs better than the FCI based on the VAR impulse responses in out-of-sample test.

Table7 R-squared and RMSE to first-order autoregressive model

AR(1)	Lag CPI	Lag CPI +FCI(SVR1)	Lag CPI +FCI(VAR)
R-squared	0.798	**0.938**	0.871
RMSE	0.294	**0.163**	0.236

In addition, when estimate VAR model, we need to consume many degree of freedom to obtain a number of parameters. Because of building FCI index of small sample size in this paper, only 54 sample points. However, if select more than four time lag, VAR models cannot be used. By contrast, SVR still makes a good simulation of economic operation process to capture financial status information of a longer lag period in smaller sample data sets, (lag 6 in this paper).

5. Conclusion

This paper proposes a new application of data mining method to construct Financial Conditions Index for forecasting CPI. We believe that the relationship between financial indicators and future CPI is more complex and non-linear, instead of linear assumption in traditional econometric. Thus, we take advantage of support vector regression algorithm using a nonlinear mapping from original financial data spaces into high dimension space, in which it constructs a linear regression function to express relationship of financial indicators and CPI simply. Actually, support vector regression also performs good forecasting using a small sample dataset. That is to say, support vector regression is a robust tool to predict CPI through using financial indicators. So, we introduce SVR to build the CPI forecasting model and make use of weight vectors to calculate weight of FCI.

In the experiment part, we make use of CPI (year on year) as the indicator of inflation, and 8 financial indicators include lagged CPI and indicators of monetary market, stock market and real estate market. We use Chinese real monthly data (from January 2006 to December 2010) to build three SVR prediction models with different lag terms. In a result, the FC(SVRs) are all leading CPI about six months around.

Furthermore, FCI (SVRs) have shown more stability comparing to traditional methods based VAR impulse response analysis. That is to say, FCI (SVRs) are more accurate, whether in sample test or out of sample test. The comparing results verify that using SVR method to construct FCI is effective. And it is worthy to apply SVR in course of financial composite index construction.

Additionally, our FCI (SVRS) are made of three FCI (SVRs) which use data of lagged 1 period, lagged 3 periods and lagged 6 periods, respectively. Actually, it does not consider the time-varying weights for a single FCI (SVR). That is to say we assume that the effect of variables on future inflation is not change with time, which does not meet the actual economic performance. Instead, the weights of our FCI (SVRs) are changing over time, which indicates lagged financial indicators impact future CPI with time-varying. Moreover, a new FCI (SVR) model which contains a time series has been posed by virtue of multiple-output SVR. Actually, a time series variables are used to predict a time series target variables. As a result, we can get rid of the problem which is mentioned above partly, in short, using the time-varying outputs instead of the time-varying weights. Such consideration will be our continue research points.

In addition to this, according to Y Shi et al.(2011)[26], we will try to use multiple criteria programming method to construct FCI, as an alternative. In this way, we can make a comparation on the effects of different optimization based data mining methods in the course of our topic. Besides, as this paper does not focus on the indicators selection, some work can be carried on in this topic. In this course, we can take advantage of knowledge management [27][28]. We believe an improvement will be gotten in that case.

Acknowledgement

This work is partially supported by National Science Foundation of China #70921061 , #71110107026 and #71071151, and by the CAS/SAFEA International Partnership Program for Creative Research Teams.

Reference

[1] Jan Hatzius, Peter Hooper, Frederic S. Mishkin, Kermit L. Schoenholtz, Mark W. Watson, Financial Conditions Indexes: A fresh look after the financial crisis. NBER Working Paper No. 16150 Issued in July 2010.
[2] V.N. Vapnik, The nature of statistical learning theory, Springer, New York, 2000:17-21.
[3] J Li, J Liu, W Xu and Yong Shi, Support Vector Machines Approach to Credit Assessment. Computational Science-ICCS2004.3039:892-899.
[4] Jingli Liu, Jianping Li, Weixuan Xu, Yong Shi, A Weighted Lq Adaptive Least Squares Support Vector Machine Classifiers-Robust and Sparse Approximation. Expert Systems With Applications. 2011, 38(3):2253-2259.
[5] Jianping Li, Zhenyu Chen , Liwei Wei, Weixuan Xu and Gang Kou, Feather Selection via Least Squares Support Feature Machine. International Journal of Information Technology & Decision Making. 2007, 6(4):671-686.
[6] Zhenyu Chen, Jianping Li, Liwei Wei, A Multiple Kernel Support Vector Machine Scheme for Feature Selection and Rule Extraction from Gene Expression Data of Cancer Tissue. Artificial Intelligence in Medicine. 2007, 41(2):161-175.
[7] Liwei Wei, Zhenyu Chen, Jianping Li, Evolution Strategies Based Adaptive Lp LS-SVM. Information Sciences. 2011, 181(14) :3000–3016.
[8] Kyoung-jae Kim, Financial time series forecasting using support vector machines. Neurocomputing. 2003(9),55(1-2):307-319.
[9] Theodore B. Trafalis, Huseyin Ince, Support vector machine for regression and application to financial forecasting. Neural Networks (IJCNN) , Proceedings of the IEEE-INNS-ENNS International Joint Conference on, 2000: 348-353
[10] M.Friedman, The Case for Flexible Exchange Rates. In: Essays in Positive Economics. The University of Chicago Press, Chicago, 1953:157-203.
[11] David G Mayes, Financial Conditions Indexes. Bank of Finland Working Paper, No.17, 2001(9).
[12] Charles Goodhart and Boris Hofmann, Asset Prices and the conduct of monetary policy. Working paper of Conference on Asset Market and Monetary Policy, 2001.
[13] Céline Gauthier and Christopher Graham,Ying Liu, Financial Conditions Indices for Canada. Bank of Canada Working paper No. 2004-2.
[14] G Dai, J,Zhang, Study on Finance Condition Index and Asset Price Channel of Monetary Policy in China. Journal of Financial and Economics, 2009(7):52-62.
[15] H. Hashem Pesarana, Yongcheol Shin, Generalized impulse response analysis in linear multivariate models. Economics letters. 1998(1), 58(1): 17–29.
[16] Steve R. Gunn, Support Vector Machines for Classification and Regression. ISIS technical report, 1998.
[17] V.N. Vapnik, The nature of statistical learning theory, Springer, New York, 2000:181-216.
[18] N Deng, Y Tian, Support vector mechine. Science Press, 2009 :64-109.(in Chinese)
[19] K Beaton, R Lalonde, C Luu, A financial conditions index for the United States. Bank of Canada research, 2009(11).
[20] Michael Holz, A Financial Conditions Index as indicator for monetary policy in times of low, stable inflation and high financial market volatility. The 9th workshop of Macroeconomics and Macroeconomic Policies, 2005.
[21] Jan Hatzius,Peter Hooper,Frederic S. Mishkin, Financial conditions indexes: a fresh look after the financial crisis. NBER Working Paper No. 16150 Issued in July 2010.
[22] B Feng, G Wang, Monetary police and financial conditions indices: a VAR analysis. Journal of Quantitative & Technical Economics, 2006(11):142-150.(in Chinese)
[23] Goldman Sachs, Euroland financial conditions too tight. European Weekly Analyst. 2003(6).
[24] Chi-Jie Lu, Tian-Shyug Lee, Chih-Chou Chiu, Financial time series forecasting using independent component analysis and support vector regression. Decision Support Systems. 2009(5),47(2):115-125.
[25] Craig Hiemstra, Jonathan D.Jones, Testing for linear and Nonliear Granger Causlity in the Stock Price Volume Relation. Journal of Finance. 1994,29(5):1941-1946.
[26] Y Shi, Y Tian, G Kou, Y Peng and J Li, Optimization Based Data Mining: Theory and Applications. Advanced Information and Knowledge Processing. Springer,2011:119-132.
[27] Lingling Zhang, Jun Li, Yong Shi and Xiaohui Liu, Foundations of Intelligent Knowledge Management, Human System Management. 2009,28 (4):145-161.
[28] Guangli Nie, Lingling Zhang,Ying Liu, Xiuyu Zheng,Yong Shi, Decision Analysis of Data Mining Project Based on Bayesian Risk. Expert Systems With Applications. 2009, 40:4589-4594.

Appendix

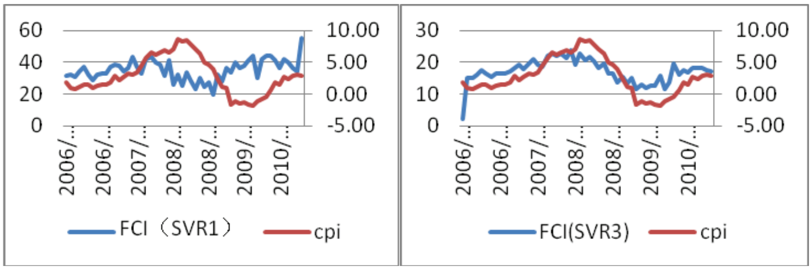

Figure 3 (a) FCI (SVR1) and CPI; (b) FCI (SVR3) and CPI

Table8 The dynamic correlations of the FCI (SVRs) with future inflation, with the respective lead displayed in first line.

	1	2	3	4	5	6	7	8
FCI (SVR1)	0. 4342	0. 5415	0. 6239	0. 6831	**0. 7016**	0. 6516	0. 568	0. 47
FCI (SVR2)	0. 1444	**0. 1459**	0. 1439	0. 0989	0. 0673	0. 0527	0. 0221	0. 0281
FCI (SVR3)	**0. 6021**	0. 5067	0. 3919	0. 2895	0. 1782	0. 0697	−0. 0418	−0. 1544

Table 9 Unit root tests for level variables

	CPI	M2	CSI300	ERI	HPI	RSR	SMC
ADF test	-1.417	-2.660***	-6.450***	-1.252	-4.484***	-1.856*	-8.823***
PP test	-1.543	-1.700*	-6.711***	-1.210	-1.853*	-1.898*	-8.451***

Note: The table displays for each variables the Augmented Dickey-Fuller and then the Phillips-Perron test statistic based on regressions with 3 lagged differences,** and *** indicate rejection of the unit root hypothesis at the 10%, 5% and 1% level respectively. The respective critical values are –1.62, -1.94 and –2.58 (MacKinnon, 1991).*

Table10 Unit root tests for first difference variales

	CPI	M2	CSI300	ERI	HPI	RSR	SMC
ADF test	-1.981**	-2.272**	-8.426***	-7.981***	-2.021**	-5.561**	-8.107***
PP test	-6.44***	-6.656***	-38.837***	-7.990***	-2.084**	-5.721**	-34.664***

Note: The table displays for each variables the Augmented Dickey-Fuller and then the Phillips-Perron test statistic based on regressions with 3 lagged differences,** and *** indicate rejection of the unit root hypothesis at the 10%, 5% and 1% level respectively. The respective critical values are –1.62, -1.94 and –2.58 (MacKinnon, 1991).*

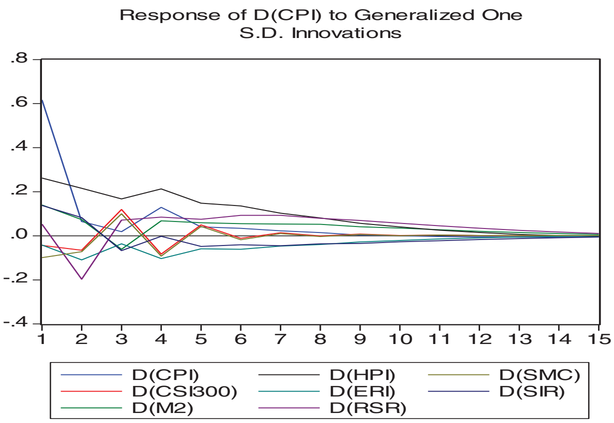

Figure 4 impulse response function figure results